

Probleme der befragungsbasierten Messung von Faktenwissen

Johann, David

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Johann, D. (2008). Probleme der befragungsbasierten Messung von Faktenwissen. *Sozialwissenschaften und Berufspraxis*, 31(1), 53-65. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-44609>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Probleme der befragungsbasierten Messung von Faktenwissen

David Johann

1 Einleitung

In den letzten Jahren ist nicht nur eine beinahe inflationäre (Quiz-)Begeisterung für Wissensfragen im nichtakademischen Bereich zu beobachten, welche sich insbesondere in den Medien niederschlägt. Auch akademische Wissensmessungen haben an Bedeutung gewonnen, so z. B. in der Jugend- und Medienforschung (Broderick 1992; Rothe 1993; Ingrisch 1997; Westle 2006) und in der Wahl- und Einstellungsforschung (Horstmann 1991; Schmitt-Beck 1993; Vetter/Maier 2005; Westle 2005). Der vielfach sorglose Umgang mit Wissensindikatoren verweist auf die Notwendigkeit, sich eingehender mit der Problematik der Wissensmessung zu befassen.

Der vorliegende Artikel beschäftigt sich mit der befragungsbasierten Messung von objektivem Wissen. Im Gegensatz zu subjektivem Wissen, welches dann vorliegt, wenn ein Individuum glaubt etwas zu wissen, unabhängig davon, ob es richtig oder falsch ist, wird objektives Wissen als faktisch zutreffendes Wissen (eben: Faktenwissen) definiert. Faktenwissen wird somit in Anlehnung an Delli-Carpini und Keeter (1996) von solchen Kognitionen abgegrenzt, die falsch sind oder nicht hinreichend getestet werden können. Mit objektivem Wissen sind darüber hinaus Kenntnisse gemeint, die im Langzeitgedächtnis gespeichert sind und sich von solchen unterscheiden, die nicht hinreichend verarbeitet oder lediglich im Kurzzeitgedächtnis gespeichert wurden. Exemplarisch liegt das Augenmerk im Nachstehenden auf dem Bereich des politischen Wissens. Die dargestellte Problematik der befragungsbasierten Messung von Wissen ist aber durchaus auch übertragbar auf andere Wissensbereiche.

Es gibt einige Divergenzen darüber, wie Wissen methodisch adäquat zu messen ist (Luskin 1987; 1990; Bennett 1988; Delli-Carpini/Keeter 1993; 1996; Lupia/McCubbins 1998; Mondak 2001). Kontrovers diskutiert wird die Instrumentenkonstruktion und dabei insbesondere die Frage, ob die Bekundung fehlenden Wissens er- oder entmutigt werden soll. Auf diese Kontroverse wird im Folgenden genauso eingegangen wie auf Operationalisierungsprobleme, die mit dem Schwierigkeitsgrad und der Trennschärfe von Fragen zu tun haben. Thematisiert werden überdies Probleme, die damit verbunden sind, dass herkömmliche Surveys nur unzureichend motivieren, Informationen abzurufen. Ziel des Artikels ist es, die Schwierigkeiten bei der Konstruktion geeigneter Instrumente zur Wissensmessung aufzuzeigen, die Auswirkungen divergierender Methoden darzulegen und den Sinn für zweckmäßige Wissensmessungen zu schärfen.

2 Schwierigkeiten bei der Messung von Faktenwissen

2.1 Frageform

In der Auseinandersetzung um das adäquate Vorgehen zur Messung von Faktenwissen spielt die Form der Frage eine zentrale Rolle. Zunächst ist nach der Art der Antwortvorgabe zwischen offenen und geschlossenen Fragen zu unterscheiden. Offene Fragen lassen dem Befragten Spielraum in der Beantwortung, sind jedoch anspruchsvoller, da die Antworten selbst formuliert werden müssen, fördern die Berücksichtigung von Meinungen und Einstellungen und sind darüber hinaus in der Implementation zeitaufwendiger. Geschlossene Fragen lassen dem Respondenten wenig Spielraum, sind schneller zu implementieren und eindeutiger zu analysieren.

Bei geschlossenen Fragen ist zwischen verschiedenen Multiple-Choice-Formaten zu unterscheiden (Haladyna 1994), von denen zur Messung von politischem Faktenwissen vor allem drei Anwendung finden: Vergleichsweise selten werden Zu- und Umordnungsaufgaben („Matching“) eingesetzt, bei denen z. B. politisches Personal den entsprechenden Parteien zugeordnet werden soll. Am gängigsten sind einfache Richtig-Falsch-Fragen sowie konventionelle Multiple-Choice-Aufgaben, welche aus einem Stamm, einer richtigen Antwortvorgabe und diversen falschen Antwortvorgaben (den „Distraktoren“) bestehen.

Ein grundsätzliches Problem der genannten Frageformen besteht darin, dass die Antworten auf die gestellten Fragen nicht wirklich das aus dem Gedächtnis abgerufene Wissen darstellen bzw. den Antworten Phänomene zugrunde liegen, die gar nicht gemessen werden sollen. Bei der Beantwortung von Faktenfragen sind sogenannte „Cues“ (Hilfestellungen) in der Fragestellung, die eine Ableitung der richtigen Antwort zulassen, oder sogar Einstellungen der Respondenten nicht zu unterschätzen (Nadeau/Niemi 1995). Bei konventionellen Multiple-Choice-Aufgaben kommt hinzu, dass nicht nur durch die Fragestellung selbst, sondern auch durch die Antwortvorgaben ungewollt Hilfestellungen gegeben werden können. Darüber hinaus ist bei dieser Art von Aufgaben strikt darauf zu achten, dass die Antwortalternativen nicht intrinsisch mehrdeutig sind. Damit die korrekte Beantwortung einer Frage nicht zu einem Puzzle wird, dürfen sich die Antwortvorgaben nicht nur in Feinheiten der Formulierung unterscheiden. Ein Indiz für eine intrinsisch mehrdeutige Aufgabe ist eine geringe Trennschärfe.

Wenn andere Dimensionen als das tatsächliche Wissen abgefragt werden, ergibt sich das Problem mangelnder Validität des Tests und mangelnder Aussagekraft über das objektive Wissen. Die Verzerrungen sind daher so gut es geht zu kontrollieren und unerwünschte Verhaltensweisen zu reduzieren. Um den Einfluss von Einstellungen und Ableitungen mit Hilfe von Cues zu verhindern oder wenigstens einzuschränken, sollten über die Wortwahl hinaus auch Einleitungen und Instruktionen Anwendung finden (Nadeau/Niemi 1995).

2.2 Antwortoptionen und Rate-Neigung

Doch auch Aufgaben mit geschlossenen Antwortvorgaben, welche sorgfältig formuliert sind, stellen Forscher vor zwei Probleme, die eng miteinander verbunden sind: Erstens

muss entschieden werden, ob eine „weiß-nicht“-Option explizit vorgegeben werden soll und wie mit ihr analytisch umzugehen ist. Zweitens besteht die Gefahr von Antwortverzerrungen aufgrund sozialer Erwünschtheit. Im Kontext von Wissensmessungen kann davon ausgegangen werden, dass die meisten Respondenten ungern Nichtwissen einräumen (Dickmann 1995). Stattdessen beantworten viele Respondenten die Wissensfragen durch Raten. Der These, dass für jegliches richtige Raten wenigstens ein minimaler Teil an Wissen vorhanden sein muss und deshalb ratende Respondenten als wenig problematisch angesehen werden können (Rothe 1993), wird hier nicht gefolgt. Durch mehr oder weniger zufällige Antworten können Verzerrungen entstehen. Da diese nicht ausgeblendet werden dürfen, ist es wichtig, sich eingehender mit der Problematik des Ratens auseinanderzusetzen.

Zunächst einmal sollte man sich mit der Anzahl der vorgegebenen Antwortmöglichkeiten beschäftigen. Bei nur zwei Antwortmöglichkeiten liegt die Wahrscheinlichkeit, richtig zu raten, mit $p=0,5$ sehr hoch. Insofern sind einfache Richtig-Falsch-Fragen problematisch, auch wenn sie den Vorteil aufweisen, dass sie weniger Zeit als Multiple-Choice-Aufgaben beanspruchen, d. h. vom Befragten schneller beantwortet und somit in der gleichen Befragungszeit mehr Fragen gestellt werden können. Bei Multiple-Choice-Aufgaben mit mehr als zwei Antwortalternativen ist die Ratewahrscheinlichkeit geringer. Daher sind sie Richtig-Falsch-Fragen vorzuziehen. Die Empfehlungen über die günstigste Anzahl an Antwortvorgaben schwanken zwischen drei und fünf (Bruno/Dirkzwager 1995; Schelten 1997; Rogers/Harley 1999; Mondak 2001; Rodriguez 2005). Wie viele Antwortalternativen tatsächlich vorgegeben werden, hängt auch davon ab, wie schwer oder leicht es fällt, plausible Antwortalternativen zu formulieren. Es mag sinnvoll sein, fünf Alternativen zu entwickeln, wenn man vier brauchbare Distraktoren konstruieren kann. Es ist aber günstiger, nur drei Alternativen zu verwenden, wenn man nicht mehr als zwei gute Distraktoren findet.

Häufige Praxis bei der Messung von Faktenwissen ist es, bei der Auswertung einen simplen Zählindex aus den richtigen Antworten zu bilden. Die falschen und die „weiß nicht“-Antworten werden als Nichtwissen gewertet (Mondak 1999). Dieses Vorgehen unterstellt, dass Wissen eine diskrete Variable ist, d. h. es gibt nur die Alternativen wissend/unwissend und keine Zwischenstufen. Daraus resultieren zunächst zwei Komplikationen: Zum einen wird nicht zwischen falsch Informierten und Uninformierten unterschieden. Während Uninformierte keine Informationen erhalten und gespeichert haben, waren falsch Informierte zumindest irgendwelchen Informationen ausgesetzt. Sie haben entweder falsche Informationen aufgenommen oder ihnen ist bei der Verarbeitung der Information oder der Erinnerung daran ein Fehler unterlaufen. Zum anderen weisen viele Respondenten bei Wissensfragen durchaus ein Teilwissen auf. Unterscheiden lassen sich demnach mit Mondak (1999; 2001) voll Informierte, teils Informierte, falsch Informierte und Uninformierte. Wenn angenommen wird, dass Wissenstests ausschließlich Wissen messen, dann impliziert dies, dass die Uninformierten mit „weiß nicht“ antworten, die falsch Informierten falsch und die Vollinformierten richtig. Ein Problem stellen die Teilinformierten dar. Es ist zu vermuten, dass ein Teil dieser Personen per Zufall richtig, ein anderer Teil per Zufall falsch und ein dritter Teil mit „weiß nicht“ antwortet, wobei unterstellt werden kann, dass die Wahrscheinlichkeit einer richtigen Antwort mit dem Umfang des Teilwissens steigt. Dies ist z. B. der Fall, wenn bei Multiple-Choice-Fragen seitens des Respondenten ein oder mehrere Antwortvorgaben ausgeschlossen werden können (Mondak 1999; 2001). Dies hat

auch zur Folge, dass zwei Formen des Ratens zu unterscheiden sind: rein zufälliges Raten („wild guessing“) und Raten auf der Grundlage partiellen Wissens („educated guessing“) (Nadeau/Niemi 1995).

Darüber hinaus ist das Vorgehen, falsche Antworten und „weiß nicht“-Antworten zu Nichtwissen zusammenzufassen, aus einem weiteren Grund problematisch: Die Wahl letztgenannter Option kann durchaus auch systematische Beziehungen zu anderen Attributen der Befragten aufweisen. So diskutieren sowohl Francis und Busch (1975) als auch Kenski und Jamieson (2000) den Zusammenhang zwischen Geschlecht und Antwortverhalten. Sie zeigen, dass Frauen wesentlich öfter als Männer dazu neigen, eine substantielle Antwort zu verweigern und zu sagen, dass sie etwas nicht wissen, obwohl sie im Fall eines Antwortzwangs etwa denselben Wissensstand aufweisen. Die psychometrische Forschung zeigt, dass es große individuelle Unterschiede in der Wahrscheinlichkeit zu raten gibt (Stanley/Hopkins 1992; Cunningham 1986). Insofern verwundert es nicht, dass unterschiedliches Antwortverhalten trotz gleicher Informationslage vor allem auf die individuell unterschiedlich ausgeprägte Rate-Neigung zurückgeführt wird. Diese Rate-Neigung setzt sich zusammen aus Persönlichkeitsfaktoren wie dem Selbstbewusstsein (Sheriff/Boomer 1954; Stanley/Hopkins 1972; Hirschfeld et al. 1995; Casey et al. 1997), der Risikofreude (Cronbach 1946; Sheriffs/Boomer 1954; Slakter 1969; Ben-Shakkar/Sinai 1991) und der Wettbewerbsorientierung (Hirschfeld et al. 1995). Personen mit großem Selbstvertrauen, geringer Risikoaversion und starker Wettbewerbsorientierung neigen eher zum Raten und verzichten folglich auf „weiß nicht“-Antworten als solche mit konträren Persönlichkeitsmerkmalen, aber gleichem Wissensstand. Da Respondenten mit dem gleichen Wissensstand auf Grund unterschiedlicher Rate-Neigung unterschiedliche Ergebnisse im Wissenstest erzielen, kann die differierende Rate-Neigung als unerwünschter Response Set betrachtet werden (Mondak 1999, 2001; Mondak/Creel Davis 2001). Werden unterschiedliche Rate-Neigungen zwischen Männern und Frauen als starker Einflussfaktor des Gender-Gaps erkannt, wie von Mondak und Anderson (2004) für die USA geschehen, so ist ein Teil des Gender-Gaps ein Artefakt. Gleiches gilt natürlich auch für Wissensunterschiede zwischen anderen Bevölkerungsgruppen, sollte sich eine divergierende Rate-Neigung als starker Prädiktor identifizieren lassen.

Die dargestellte Problematik der Rate-Neigung führt zu einer Frage, die seit einigen Jahren in der amerikanischen politikwissenschaftlichen Literatur kontrovers diskutiert wird: Ist die Bekundung fehlenden Wissens zu forcieren, um die Rate-Neigung einzudämmen, oder soll umgekehrt zu inhaltlichen Antworten gedrängt werden, um alle Respondenten im Fall von Unsicherheit gleichermaßen zum Raten zu animieren (Delli Carpini/Keeter 1996; Mondak 2001).

Lange war es in politikwissenschaftlichen Surveys üblich, die Respondenten darin zu bestärken, Nichtwissen einzuräumen. Vorrangiges Ziel war dabei, das Raten zu verhindern oder zumindest zu reduzieren. Darüber hinaus ging man davon aus, dass schlecht informierte Respondenten dadurch nicht beunruhigt oder entmutigt würden (Mondak/Creel Davis 2001). Um dieses Verhalten zu motivieren, wurde explizit eine „weiß nicht“-Option vorgegeben und/oder in der Einleitung und Hinführung zum Test aufgefordert, Nichtwissen bei Unsicherheit einzugestehen. Außerdem war die Wortwahl in den Fragen indirekt und nicht direkt („Wissen Sie vielleicht [...]?“). Und schließlich wurden die Interviewer dazu angehalten, nicht nachzufragen, wenn sich die Respondenten nicht sicher waren (Schumann/Presser 1981; Delli Carpini/Keeter 1993).

Psychologische, lerntheoretische Befunde deuten darauf hin, dass mit dem beschriebenen Vorgehen die Rate-Neigung nicht hinreichend verringert wird, sondern eher Mut belohnt und Unsicherheit bestraft wird, die Befunde also durch kognitionsfremde Elemente verzerrt werden. Mondak und Creel Davis (2001) betonen, dass die Ergebnisse zwar im Sinne der Reliabilität zuverlässiger würden, die Validität hingegen nur erhöht würde, wenn zwei sehr unwahrscheinliche Annahmen zuträfen: Nur die Uninformierten müssten von substantiellen Antworten abgehalten werden, und alle Uninformierten müssten substantielle Antworten verweigern. Entgegen der Vorgehensweise, „weiß nicht“-Antworten zu ermutigen, setzen sich Mondak (1999, 2001), Mondak und Creel Davis (2001) sowie Krosnick et al. (2002) daher dafür ein, zu substantiellem Antworten zu ermutigen. Diese in der Psychologie und Pädagogik anerkannte Praxis bringe zwar Zufallsfehler mit sich, welche die Reliabilität zwischen den Items reduziere, habe aber den Vorteil, dass die Ergebnisse nicht durch unterschiedliche Neigungen zum Raten verzerrt würden, Wissen nicht verborgen bliebe und die Befragten ihr Wissen optimal darstellten, was zu valideren Erkenntnissen über den Wissensstand führe (Cronbach 1946; Slakter 1969).

Der Einfluss der individuellen Rate-Neigung kann nach Mondak (1999) und Mondak und Creel Davis (2001) eliminiert werden, indem Multiple-Choice-Fragen offenen Fragen vorgezogen werden, die „weiß nicht“-Option nicht explizit zugelassen wird, die Respondenten in der Einleitung ermutigt werden, substantielle Antworten zu geben, sowie direkte Fragen an die Befragten gestellt und keine abmildernden Phrasen eingesetzt werden. Darüber hinaus sollten die Interviewer die Respondenten zu substantiellem Antworten drängen, wenn sie Unsicherheit zeigen. Wenn doch „weiß nicht“-Ausprägungen aufträten, sollten diese per Zufall den Antwortmöglichkeiten zugeordnet werden. Außerdem solle man darauf verzichten, bei der Indexbildung für falsche Antworten Punkte abzuziehen. Jedoch sind auch die hier gemachten Vorschläge mit Vorsicht zu genießen. Die Prozedur, „weiß nicht“-Antworten zufällig substantiellen Antworten zuzuordnen, geht davon aus, dass die Respondenten, welche die „weiß nicht“-Option gewählt haben, rein zufällig geraten hätten, wenn sie zu einer substantiellen Antwort gedrängt worden wären. Möglich scheint aber, dass systematisch bedingt, bspw. aufgrund eines größeren Selbstbewusstseins oder aufgrund von Teilwissen, mehr richtige oder mehr falsche Antworten gegeben werden.

2.3 Itemschwierigkeit und Trennschärfe

Die genannten Probleme der Instrumentenkonstruktion sind nicht die einzigen Operationalisierungsprobleme. In die Entscheidung für oder gegen eine Frage fließen außerdem der Schwierigkeitsgrad und die Trennschärfe ein. Mit Schwierigkeitsgrad ist gemeint, wie schwer oder leicht es ist, eine Test- bzw. Wissensaufgabe zu lösen. Zur Bestimmung der Schwierigkeit einer Aufgabe wird ein Schwierigkeitsindex gebildet. Er gibt an, zu welchem Anteil die Respondenten die Frage richtig beantworten. Der Trennschärfekoeffizient eines Items ist dagegen der Korrelationskoeffizient aus dem Testwert dieses Items und dem Gesamtwert über alle Items. Er zeigt an, wie gut das gesamte Testergebnis durch das einzelne Item repräsentiert wird. Ebenso wie der Schwierigkeitsindex liefert der Trennschärfekoeffizient einen Hinweis, wie gut eine Frage in der Lage ist, zwischen viel und wenig Wissen in einem Test zu unterscheiden (Schelten 1997; Bortz/Döring 2002).

Weder Fragen, die von nahezu allen, noch solche, die von beinahe niemandem richtig beantwortet werden, tragen zur Differenzierung viel und wenig wissender Personen bei. Zu

leichte Fragen bergen überdies die Gefahr, dass die Respondenten sich nicht ernst genommen fühlen oder langweilen, während zu schwierige Fragen zu einer Überforderung und/oder Verärgerung der Respondenten führen können. Mit Ausnahme einiger leichter Fragen zum Einstieg sowie gegebenenfalls bei längeren Fragebögen zur Aufrechterhaltung der Motivation wird daher i. d. R. empfohlen, darauf zu achten, sowohl extrem leichte als auch extrem schwere Fragen zu vermeiden. Während Delli-Carpini und Keeter (1993) Aufgaben mit einem Schwierigkeitsgrad zwischen $p=0,3$ und $p=0,7$ (also einem Anteil richtiger Antworten zwischen 30 und 70 Prozent) empfehlen, bezeichnet Schelten (1997) Schwierigkeitsgrade von $p=0,35$ bis $p=0,75$ als annehmbar. Dass der Schwierigkeitsgrad einer Wissensfrage in diesem Bereich liegt, sei eine notwendige, nicht aber eine hinreichende Bedingung für die Trennschärfe einer Frage. Der Trennschärfekoeffizient kann Werte zwischen -1 und 1 annehmen. Negativ trennscharf und damit äußerst problematisch ist eine Frage dann, wenn ansonsten wenig wissende Respondenten sie zu hohen Anteilen richtig beantworten, während die in der Regel viel wissenden Befragten sie falsch beantworten. Weist ein Item negative Trennschärfe auf, generiert sie Informationen, die nicht mit dem Gesamtergebnis übereinstimmen. Daher ist es nicht einfach aus dem Test zu entfernen, sondern es ist eine Analyse der ganzen Testaufgabe sowie idealtypisch, aber im Rahmen vieler Projekte nicht zu leisten, eine zusätzliche Befragung der getesteten Personen ratsam, da so Auskünfte gesammelt werden können, wie die Antwortverteilung zustande kam. Empfohlen werden hohe Trennschärfen: Während Aufgaben mit einem Trennschärfekoeffizienten von unter 0,3 als wenig geeignet bzw. unbrauchbar charakterisiert werden, seien Aufgaben mit einem Trennschärfekoeffizienten zwischen 0,3 und 0,5 als mittelmäßig und damit akzeptabel und Aufgaben mit Werten größer 0,5 als hoch bzw. ausgezeichnet anzusehen (Weise 1975; Schelten 1997).

Auch bei dieser auf die Klassische Testtheorie zurückgehenden Analyse der Items ist auf Probleme hinzuweisen. Die berechneten Eigenschaften der Testaufgaben und die Fähigkeiten der Respondenten sind miteinander verbunden. Die Testleistungen sind also nicht ausschließlich durch die Qualität der ausgewählten Items im Test bestimmt, sondern die Ergebnisse einer Person sind letztlich auch populations- bzw. stichprobenabhängig.¹ Als Folge weist ein klassisch konstruierter Wissenstest eine unterschiedliche Güte für unterschiedliche Subgruppen der Stichprobe auf. So kann der Schwierigkeitsgrad mit Betroffenheiten in der Bevölkerung korrespondieren. Z. B. ist es denkbar, dass ältere Respondenten eher über rentenpolitische Fragen Bescheid wissen als jüngere, während sich letztere besser mit Fragen der Bildungs- oder Umweltpolitik auskennen. Im Endeffekt bedeutet dies auch, dass nicht von vornherein von einer „generalistischen“ Wissensstruktur ausgegangen werden kann, sondern die Möglichkeit der Multidimensionalität von Wissen zu prüfen ist, soll ein Survey Informationen über Wissensstände und -strukturen der Allgemeinbevölkerung liefern. Darüber hinaus ist strikt darauf zu achten, dass die Merkmale der Stichprobe, an der der Wissenstest (mittels eines Pretests) entwickelt wird, den Merkmalen der interessierenden Gruppe in größtmöglichem Umfang entsprechen – was aber dann schwer umsetzbar sein dürfte, wenn noch keine oder nur wenige Informationen über die

1 Schwächen der Klassischen Testtheorie wie die Abhängigkeit der Schwierigkeit und der Trennschärfe von der untersuchten Personenstichprobe versprechen Modelle der modernen Testtheorie (Item Response Theorie) zu lösen. Inwieweit in diesen Modellen tatsächlich von Stichprobenunabhängigkeit ausgegangen werden kann, ist aber umstritten (Lippert et al. 1977).

Wissensstände und -strukturen sowie die ihnen zugrunde liegenden Einflussfaktoren in der angestrebten Grundgesamtheit vorliegen.

Ein weiterer Punkt ist anzusprechen: Breiter Konsens ist, dass Aufgaben mit mittlerer Schwierigkeit und hoher Trennschärfe zur Leistungsdifferenzierung beitragen und solchen mit geringem bzw. hohem Schwierigkeitsgrad sowie niedriger Trennschärfe vorzuziehen sind. Diese nach einer Norm sortierende Testtheorie ist konstruiert worden um (Wissens-)Unterschiede zwischen Testpersonen zu erfassen. Sie beseitigt aber Items, die in einer Stichprobe alle in einem bestimmten Sinne beantwortet werden, also keine Streuung aufweisen. Für einige Fragestellungen kann dies unangebracht sein, denn Items, die von allen bzw. niemandem richtig beantwortet werden können und aus Sicht der Testtheorie zu eliminieren sind, kennzeichnen die Bereiche am besten, über die (annähernd) jeder bzw. keiner Bescheid weiß.

2.4 Motivation

In dem Bemühen, die Ausschöpfungsquote zu erhöhen, werden in Studien zunehmend sogenannte „Incentives“ eingesetzt. Incentives sind monetäre und nicht-monetäre Anreize, welche die Respondenten veranlassen sollen, sich an der Umfrage zu beteiligen. Prominente Beispiele für Studien, bei denen mit Incentives gearbeitet wurde, sind im angloamerikanischen Raum die American National Election Study 2000 und Surveys, die vom Office for National Statistics (ONS) in Großbritannien durchgeführt werden, wie der Omnibus Survey, der Family Resources Survey, der Expenditure and Food Survey und der General Household Survey. Eingesetzt wurden sowohl nichtmonetäre Anreize wie Stifte oder Briefmarken, die mit dem Kontaktschreiben versandt wurden, als auch monetäre Anreize in Form von kleinen Geldbeträgen, welche die Respondenten erhielten, wenn sie die Fragebögen komplett ausfüllten (z. B. 20 Dollar bei der American National Election Study 2000 bzw. 5 bis 10 Pfund beim Expenditure and Food Survey). Auch in Deutschland wird mit Incentives experimentiert. So erhielten beim Allbus/ISSP 2002 alle Befragten für die Teilnahme eine Sondermünze im Wert von 10 Euro (Blohm et al. 2003). In aller Regel ist der Einsatz von Incentives effektiv. Er führt zu einer Erhöhung der Ausschöpfungsquote, und zwar sowohl bei postalischen als auch bei telefonischen sowie Face-to-Face-Surveys. Die Höhe der Incentives spielt dabei nur eine untergeordnete Rolle (Simmons/Wilmot 2004).

Nun ist das Bemühen, die Ausschöpfungsquote zu erhöhen, nicht auf Surveys begrenzt, in denen Faktenwissen erhoben wird. Der Gedanke, Anreize einzusetzen, um die Motivation der Respondenten zu fördern, ist in diesem Zusammenhang aber von besonderer Bedeutung und darüber hinaus keinesfalls nur auf die Teilnahmebereitschaft zu beschränken. So weisen Lupia und Prior (2005) darauf hin, dass durchaus auch Anreize gegeben werden sollten, die einzelnen Wissensfragen sorgfältig und mit Bedacht zu beantworten. In einem national repräsentativen Survey eingebettete Experimente zu politischem Wissen zeigten, dass es herkömmliche Surveys nicht vermögen, die moderat oder stark interessierten Bevölkerungsteile dazu zu bringen, sich genügend anzustrengen. Monetäre Anreize in der Form, dass für jede korrekte Antwort ein bestimmter Betrag gezahlt wurde (hier: 1 Dollar), führten aber dazu, dass der Anteil der richtigen Antworten in diesen Gruppen stieg. Dieses Ergebnis zeigt, dass die Respondenten mehr Informationen bereithalten und verarbeiten, als allgemein angenommen wird. Dementsprechend vertreten Lu-

pia und Prior die Meinung, dass eine solche Messung aussagekräftiger ist als eine konventionelle.

Die Ergebnisse unterstellen, dass herkömmliche Wissensmessungen nicht ausreichend motivieren, (politische) Information abzurufen. Es werden zu wenige Anreize gegeben, die Items gewissenhaft zu beantworten, was zu einer Unterschätzung des Wissens führt. Ließe sich feststellen, dass diese Verzerrung zeitlich konstant ist, dürften herkömmliche Wissensmessungen zumindest die Entwicklungsrichtung der Wissensstände korrekt wiedergeben, wenn auch auf einem niedrigeren Niveau. Problematisch ist aber der Befund, dass bestimmte Gruppen unmotivierter sind als andere. Die Wissensstände werden nicht in allen Bevölkerungsgruppen gleichermaßen unterschätzt. Vielmehr treten Unterschätzungen unter den politisch wenig interessierten Bevölkerungsteilen seltener auf. Herkömmliche Wissensmessungen können folglich den Eindruck erwecken, die Wissensunterschiede zwischen den wenig Interessierten und den moderat bzw. stark Interessierten seien weniger groß, als dies tatsächlich der Fall ist.

3 Diskussion

Für einige Forscher erscheint der Prozess der Beantwortung von Faktenfragen einfach: Respondenten, die die Antwort wissen, rufen die Lösung aus dem Gedächtnis ab und geben die richtige Antwort. Diejenigen dagegen, die die richtige Antwort nicht kennen, bekunden fehlendes Wissen oder geben eine falsche Antwort, wobei bekundetes fehlendes Wissen und falsche Antworten gleichermaßen als Nichtwissen gewertet werden. Bei der Beantwortung von Faktenfragen handelt es sich jedoch mitnichten um einen einfachen, sondern im Gegenteil um einen komplizierten Prozess. So werden sich bspw. einige Respondenten, die sich unsicher sind, zu einer substantiellen Antwort verpflichtet fühlen, andere dagegen nicht. Bei der Beantwortung der Fragen spielen außerdem bestimmte Reize, das Schlussfolgern und die Verfügbarkeit der Erinnerung eine Rolle. So kann es sein, dass seitens der Respondenten aus Informationsbruchstücken Heuristiken gebildet werden, die die korrekte Beantwortung einer Frage ermöglichen.

Respondenten sind keine homogene Menge. In manchen Fällen sind die Antworten sorgfältig aus einem großen Wissensvorrat abgeleitet, in anderen Fällen wird geraten: zufällig oder auf der Grundlage partiellen Wissens. Die psychometrische Forschung zeigt ferner, dass es große individuelle Unterschiede in der Wahrscheinlichkeit zu raten gibt. Wissensmessungen werden demnach auch durch Persönlichkeitsfaktoren wie das Selbstbewusstsein, die Wettbewerbsorientierung und die Risikofreude kontaminiert. Durch die explizite Ermutigung der „weiß nicht“-Option aktiviert man ein Response Set, nämlich die unterschiedliche Neigung zu raten. Solche Response Sets sind bei offenen Fragen noch wahrscheinlicher als bei Multiple-Choice-Aufgaben, weil bei letztgenannter Frageform unsichere Respondenten durch die Existenz der von ihnen vermuteten Antwort unter den Vorgaben bestärkt werden. Im Gegensatz zu vielen politikwissenschaftlichen Surveys versuchen erziehungswissenschaftliche Studien zumeist die Rate-Neigung der Respondenten auf ein identisches Niveau zu bringen, um die Validität und Reliabilität der Wissensskalen als rein kognitive Maße zu optimieren. Der Autor vertritt die Auffassung, dass dieses Prozedere in der Tat eher auf den rein kognitiven Aspekt von Wissen zielt. Der Einbezug von „weiß nicht“ nimmt hingegen Interferenzen durch Unsicherheiten in Kauf. Dass solche

Unsicherheiten auch für die Folgen von Wissen relevant sind, also bspw. für das handlungsrelevante Wissen valider sind, spricht dennoch für diese Vorgehensweise (Westle 2007).

Im Zusammenhang mit Wissensmessungen werden vielfach die Konsequenzen diskutiert, die mit mehr oder weniger Wissen verknüpft sind. Im Bereich des politischen Wissens wird bspw. aus normativ demokratietheoretischer Perspektive kritisch angemerkt, die Bevölkerung sei aufgrund des geringen Wissensstandes nicht in der Lage, politische Entscheidungen zu treffen, die für sie vorteilhaft sind (Bennett 1989; Delli-Carpini/Keeter 1996). Die Forschung zu (politischem) Wissen sollte aber immer dann mit Vorsicht genossen werden, wenn Rückschlüsse auf die Fähigkeit bzw. Unfähigkeit der Befragten gezogen werden. Herkömmliche Wissensmessungen liefern keine hinreichenden Informationen über die Kompetenz, sich Wissen zugänglich zu machen und anzueignen. Sie unterschätzen somit die Kapazität der Respondenten, informierte Entscheidungen zu treffen. Kognitionspsychologen unterscheiden nach Art der Gedächtnisinhalte im Langzeitgedächtnis zwischen dem deklarativen und dem prozeduralen Gedächtnis. Während im deklarativen Gedächtnis Fakten gespeichert sind, beinhaltet das prozedurale Gedächtnis die Fertigkeit zu wissen, wie und wo man (verhaltensrelevante) Informationen findet (Kandel et al. 1995). Bei herkömmlichen Wissensmessungen greifen die Respondenten auf den Teil des Gedächtnisses zurück, der als deklaratives Gedächtnis bezeichnet wird (National Research Council 1994). Im Gegensatz zum deklarativen Gedächtnis kann das prozedurale Gedächtnis nicht direkt beobachtet werden. Natürlich gibt es viele Situationen, in denen man kurzfristig, ohne Vorwarnung Wissen braucht. In anderen Situationen bleibt aber genügend Zeit, relevante Informationen zu sammeln und so begründete Entscheidungen zu treffen. Auf diese Situationen können mit traditionellen Wissenssurveys nur eingeschränkt Rückschlüsse gezogen werden (Lupia/Prior 2005).

Alles in allem kann ein zu sorgloses Vorgehen bei der Instrumentenkonstruktion schwerwiegende Auswirkungen auf die empirischen Befunde und damit die (Möglichkeiten der) Interpretation haben. Dies betrifft nicht nur die Frage des Gender-Gaps, sondern auch Unterschiede im Lebenslauf oder zwischen den Generationen sowie Zusammenhänge von Wissen und Interesse der Respondenten. Der Konstruktionsphase eines Wissenstests kommt eine zentrale Bedeutung zu. In diesem Stadium entscheidet sich, wie verlässlich die Resultate sein werden. Angesichts der Mannigfaltigkeit der mit der Konstruktion eines hochwertigen Erhebungsinstruments verknüpften Probleme spielt der Pretest bei Wissensmessungen eine besondere Rolle. Er sollte nicht an einer zu kleinen Befragtenzahl durchgeführt werden, da sonst keine hinreichenden Aussagen über Strukturprobleme der Indikatoren möglich sind. Außerdem ist unbedingt darauf zu achten, dass in Pretestverfahren weithin gängige Verzerrungen der Probandenpopulation, bspw. durch besonders befragungsbereite oder außergewöhnlich interessierte Personen, vermieden werden.

Trotz der Vielzahl an Schwierigkeiten, die mit der befragungsbasierten Messung von Faktenwissen verbunden sind, können gewissenhaft entwickelte Instrumente wichtige Informationen liefern, so z. B. zur Beantwortung der Frage, wie Wissen in der Bevölkerung strukturiert ist oder wie Wissen Einstellungen beeinflusst.

Literatur

- Ben-Shakhar, Gershon; Sinai, Yakov, 1991: Gender Difference in Multiple-Choice Tests: The Role of Differential Guessing Tendencies. In: *Journal of Educational Measurement*, Vol. 28, S. 23-35.
- Bennett, Stephen E., 1988: „Know-Nothings“ Revisited: The Meaning of Political Ignorance Today. In: *Social Science Quarterly*, Vol. 69, No. 2, S. 476-490.
- Bennett, Stephen E., 1989: Trends in Americans' Political Information, 1967-1987. In: *American Politics Quarterly*, Vol. 17, No. 4, S. 422-435.
- Blohm, Michael; Harkness, Janet; Klein, Sabine; Scholz, Evi, 2003: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2002. ZUMA Methodenbericht 12/2003.
- Bortz, Jürgen; Döring, Nicola, 2002: *Forschungsmethoden und Evaluation*. Berlin, Heidelberg, New York: Springer-Verlag.
- Broderick, Alesia, 1992: Media Use and Political Knowledge: A Comparison Between Adolescents in East and West Germany. Regensburg: Unveröffentlichte Diplomarbeit.
- Bruno, James E.; Dirkzwager, Arie, 1995: Determining the Optimal Number of Alternatives to a Multiple-Choice Test Item: An Information Theoretic Perspective. In: *Educational and Psychological Measurement*, Vol. 55, No. 6, S. 959-966.
- Casey, M. Beth; Nuttall, Ronald L.; Pezaris, Elizabeth, 1997: Mediators of Gender Differences in Mathematics College Entrance Test Scores: A Comparison of Spatial Skills with Internalized Beliefs and Anxieties. In: *Developmental Psychology*, Vol. 33, No. 4, S. 669-680.
- Cronbach, Lee J., 1946: Response Sets and Test Validity. In: *Educational and Psychological Measurement*, Vol. 6, No. 3, S. 475-494.
- Delli Carpini, Michael X.; Keeter, Scott, 1993: Measuring Political Knowledge: Putting First Things First. In: *American Journal of Political Science*, Vol. 37, No. 4, S. 1179-1206.
- Delli Carpini, Michael X.; Keeter, Scott, 1996: *What Americans Know about Politics and Why It Matters*. Yale: University Press.
- Diekmann, Andreas, 1995: *Empirische Sozialforschung*. Reinbek: Rowohlt.
- Francis, Joe D.; Busch, Lawrence, 1975: What we know about “I don't knows“. In: *Public Opinion Quarterly*, Vol. 39, No. 2, S. 207-218.
- Haladyna, Thomas M., 1994: *Developing and Validating Multiple-Choice Test Items*. Hillsdale: Lawrence Erlbaum Associates.
- Hirschfeld, Mary; Moore, Robert L.; Brown, Eleanor, 1995: Exploring the Gender Gap on the GRE Subject Test in Economics. In: *Journal of Economic Education*, Vol. 26, No. 1, S. 3-15.
- Holbrook, Allyson L.; Berent, Matthew K.; Krosnick, Jon A.; Visser, Penny S.; Boninger, David S., 2005: Attitude importance and the accumulation of attitude-relevant knowledge in memory. In: *Journal of Personality and Social Psychology*, Vol. 88, No. 5, S. 749-769.
- Horstmann, Reinhold, 1991: *Medieneinflüsse auf politisches Wissen. Zur Tragfähigkeit der Wissenskluft-Hypothese*. Wiesbaden: Deutscher Universitätsverlag.

- Ingrisch, Michaela, 1997: Politisches Wissen, politisches Interesse und politische Handlungsbereitschaft bei Jugendlichen aus den alten und neuen Bundesländern. Regensburg: Univ. Diss.
- Kandel, Eric R.; Schwartz, James H.; Jessell, Thomas M., 1995: Neurowissenschaften. Eine Einführung. Heidelberg: Spektrum Akademischer Verlag.
- Kenski, Kate; Jamieson, Kathleen Hall, 2000: The gender gap in political knowledge: Are women less knowledgeable than men about politics? In: Jamieson, Kathleen Hall (Hrsg.): Everything you think you know about politics... and why you're wrong. New York: Basic Books, S. 83-89.
- Krosnick, Jon A.; Holbrook, Allyson L.; Berent, Matthew K.; Carson, Richard T.; Hanemann, W. Michael; Kopp, Raymond J.; Mitchell, Robert Cameron; Presser, Stanley; Ruud, Paul R.; Smith, V. Kerry; Moody, Wendy R.; Green, Melanie C.; Conoway, Michael, 2002: The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfy? In: Public Opinion Quarterly, Vol. 66, No. 3, S. 371-403.
- Lippert, Ekkehard; Schneider, Paul; Wakenhut, Roland, 1977: Zur Stabilität probabilistischer Skalierungsverfahren. In: Psychologische Beiträge, Nr. 19, S. 588-599.
- Lupia, Arthur; Prior, Markus, 2005: What Citizens Know Depends on How You Ask Them: Political Knowledge and Political Learning Skills. MPRA Paper No. 103: http://mpra.ub.uni-muenchen.de/103/01/MPRA_paper_103.pdf. Stand: 15. Mai 2007.
- Lupia, Arthur; McCubbins, Mathew, 1998: The Democratic Dilemma: Can Citizens Learn What They Need to Know? Cambridge: University Press.
- Luskin, Robert C., 1987: Measuring Political Sophistication. In: American Journal of Political Science, Vol. 31, No. 4, S. 856-899.
- Luskin, Robert C., 1990: Explaining Political Sophistication. In: Political Behavior, Vol. 12, No. 4, S. 331-361.
- Mondak, Jeffrey J., 1999: Reconsidering the measurement of political knowledge. In: Political Analysis, Vol. 8, No. 1, S. 57-82.
- Mondak, Jeffrey J., 2001: Developing valid knowledge scales. In: American Journal of Political Science, Vol. 45, No. 1, S. 224-238.
- Mondak, Jeffrey J.; Creel Davis, Belinda, 2001: Asked and answered: Knowledge levels when we will not take "don't know" for an answer. In: Political Behavior, Vol. 23, No. 3, S. 199-224.
- Mondak, Jeffrey J.; Anderson, Mary R., 2004: The knowledge gap: A reexamination of gender-based differences in political knowledge. In: The Journal of Politics, Vol. 66, No. 2, S. 492-512.
- Nadeau, Richard; Niemi, Richard G., 1995: Educated guesses: The process of answering factual knowledge questions in surveys. In: Public Opinion Quarterly, Vol. 59, No. 3, S. 323-346.
- National Research Council, 1994: Transfer: Training for Performance. In: Druckman, Daniel; Bjork, Robert A. (Hrsg.), 1994: Learning, Remembering, Believing: Enhancing Human Performance. Washington: National Academy Press, S. 25-56.
- Reuband, Karl-Heinz, 2000: „Pseudo Opinions“ in Bevölkerungsumfragen. Wie die Bürger fiktive Politiker beurteilen. In: ZA-Information 46, S. 26-38.

- Rodriguez, Michael C., 2005: Three options are optimal for multiple-choice questions: A meta-analysis of 80 years of research. In: *Educational Measurement: Issues & Practice*, Vol. 24, No. 2, S. 3-13.
- Rogers, W. Todd; Harley, Dwight, 1999: An Empirical Comparison of Three-and Four-Choice Items and Tests: Susceptibility to Testwiseness and Internal Consistency Reliability. In: *Educational and Psychological Measurement*, Vol. 59., No. 2. S. 234-247.
- Rothe, Klaus, 1993: *Schüler und Politik: Eine vergleichende Untersuchung bayerischer und hessischer Gymnasialschüler*. Opladen: Leske + Budrich.
- Schelten, Andreas, 1997: *Testbeurteilung und Testerstellung: Grundlagen der Teststatistik und Testtheorie für Pädagogen und Ausbilder in der Praxis*. Stuttgart: Steiner.
- Schmitt-Beck, Rüdiger, 1993: Denn sie wissen nicht, was sie tun... Zum Verständnis des Verfahrens der Bundestagswahl bei westdeutschen und ostdeutschen Wählern. In: *Zeitschrift für Parlamentsfragen*, 24. Jg., Heft 3, S. 393-415.
- Schuman, Howard; Presser, Stanley, 1981: *Questions and Answers in Attitude Surveys*. New York: Academic Press, Inc.
- Schnell, Rainer; Hill, Paul B.; Esser, Elke, 1999: *Methoden der empirischen Sozialforschung*. München [u.a.]: Oldenbourg.
- Sheriffs, Alex C.; Boomer, Donald S., 1954: Who is Penalized by the Penalty for Guessing? In: *Journal of Educational Psychology*, Vol. 45, No. 2, S. 81-90.
- Simmons, Eleanor; Wilmot, Amanda, 2004: Incentive payments on social surveys: A literature review. In: *Survey Methodology Bulletin*, Vol. 53, No. 1, S. 1-11.
- Slakter, Malcolm J., 1969: Generality of Risk Taking on Objective Examinations. In: *Educational and Psychological Measurement*, Vol. 29, No. 1, S. 115-128.
- Stanley, Jullian C.; Hopkins, Kenneth D., 1972: *Educational and Psychological Measurement and Evaluation*. Englewood Cliffs, NJ: Prentice-Hall.
- Vetter, Angelika; Maier, Jürgen, 2005: Mittendrin statt nur dabei? Politisches Wissen, politisches Interesse und politisches Kompetenzgefühl in Deutschland, 1994-2002. In: Gabriel, Oscar W.; Falter, Jürgen W.; Rattinger, Hans (Hrsg.), 2005: *Wächst zusammen, was zusammengehört? Stabilität und Wandel politischer Einstellungen im wiedervereinigten Deutschland*. Baden-Baden: Nomos, S. 51-90.
- Westle, Bettina, 2005: Politisches Wissen und Wahlen. In: Falter, Jürgen W.; Gabriel, Oscar W.; Weißels, Bernhard (Hrsg.), 2005: *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2002*. Wiesbaden: Verlag für Sozialwissenschaften, S. 484-512.
- Westle, Bettina, 2006: Politisches Interesse, subjektive politische Kompetenz und politisches Wissen – Eine Fallstudie mit Jugendlichen im Nürnberger Raum. In: Roller, Edeltraud; Brettschneider, Frank; Deth, Jan W. v. (Hrsg.), 2006: *Jugend und Politik: „Voll Normal!“. Der Beitrag der politischen Soziologie zur Jugendforschung*. Wiesbaden: Verlag für Sozialwissenschaften, S. 209-240.
- Westle, Bettina, 2007: *Stand, Struktur und Entwicklung des politischen Wissens in der deutschen Bevölkerung*. Unveröffentlichter Projektantrag.

David Johann

Philipps-Universität Marburg

Institut für Politikwissenschaft

Politikwissenschaftliche Methoden & empirische Demokratieforschung

Wilhelm-Röpke-Straße 6 G
D-35032 Marburg
Tel.: +49 (0)6421 28-24325
E-Mail: mail@davidjohann.de

David Johann, Dipl. Soziologe, geb. 1979, studierte Soziologie, Politologie, Psychologie und Volkswirtschaftslehre an der Freien Universität Berlin. Seit 2006 ist er wissenschaftlicher Mitarbeiter am Institut für Politikwissenschaft der Philipps-Universität Marburg, Lehrinheit für politikwissenschaftliche Methoden und empirische Demokratieforschung. Seine Arbeitsschwerpunkte sind: Methoden der empirische Sozialforschung, Wahl- und Einstellungsforschung, Parteienforschung, Arbeits- und Wirtschaftssoziologie, Genderforschung.